PNAS RESEARCH ARTICLE | CHEMISTRY
BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Odd–even disparity in the population of slipped hairpins in RNA repeat sequences with implications for phase separation

Hiranmay Maity[a], Hung T. Nguyen[a], Naoto Hori[b], and D. Thirumalai[a,c,1]

Edited by Gerhard Hummer, Max Planck Institute for Biophysics, Frankfurt am Main, Germany; received January 25, 2023; accepted May 10, 2023, by Editorial Board Member J. A. McCammon

Low-complexity nucleotide repeat sequences, which are implicated in several neurological disorders, undergo liquid–liquid phase separation (LLPS) provided the number of repeat units, *n*, exceeds a critical value. Here, we establish a link between the folding landscapes of the monomers of trinucleotide repeats and their propensity to self-associate. Simulations using a coarse-grained Self-Organized Polymer (SOP) model for (CAG)$_n$ repeats in monovalent salt solutions reproduce experimentally measured melting temperatures, which are available only for small *n*. By extending the simulations to large *n*, we show that the free-energy gap, ΔG_S, between the ground state (GS) and slipped hairpin (SH) states is a predictor of aggregation propensity. The GS for even *n* is a perfect hairpin (PH), whereas it is a SH when *n* is odd. The value of ΔG_S (zero for odd *n*) is larger for even *n* than for odd *n*. As a result, the rate of dimer formation is slower in (CAG)$_{30}$ relative to (CAG)$_{31}$, thus linking ΔG_S to RNA–RNA association. The yield of the dimer decreases dramatically, compared to the wild type, in mutant sequences in which the population of the SH decreases substantially. Association between RNA chains is preceded by a transition to the SH even if the GS is a PH. The finding that the excitation spectrum—which depends on the exact sequence, *n*, and ionic conditions—is a predictor of self-association should also hold for other RNAs (mRNA for example) that undergo LLPS.

self-organized polymer model | low complexity RNA sequences | RNA–RNA association | excited states | liquid–liquid phase separation

Significance

RNA sequences, [(CAG)$_n$ and (CUG)$_n$], are linked to neurological disorders provided the repeat number *n* exceeds a threshold value. The factors determining the propensities and the mechanism of condensate formation in the low-complexity RNA sequences are unknown. Simulations using coarse-grained models of (CAG)$_n$ as a function of *n* show that for even *n*, the ground state is a perfect hairpin (PH), whereas as for odd *n*, it is a slipped hairpin (SH) with one or more CAG overhangs. The propensities and rates of self-association are inversely correlated with the free-energy gap separating the ground state and the SH. Self-association between RNA is always preceded by SH formation. Our findings explain the absence of aggregation in typical heterogeneous sequences.

The pioneering study by Eisenberg and Felsenfeld (1) showed that polyriboadenylic acid (poly rA) could undergo phase separation in which a dense phase (condensate) coexists with the dispersed (sol) phase. However, it is only recently the importance of condensate formation involving RNA molecules in a variety of contexts is starting to be appreciated (2–10). The diversity of RNA sequences and the myriad structures an isolated RNA (a self-organizing polymer) adopts make it difficult to uncover the molecular features that drive phase separation. Nevertheless, one could anticipate a few scenarios for phase separation by favorable RNA–RNA interactions. 1) RNA molecules, containing unpaired single-strand regions, may have a high propensity to associate with other RNA molecules, through complementary base pair formation, to form aggregates (11, 12). In principle, both homotypic and heterotypic interactions are possible, but in many cases, one observes predominantly homotypic interactions, especially when there is sequence diversity (4, 7). 2) The presence of a single strand in the ground state (lowest free-energy state) that could engage in Watson–Crick (WC) base pair formation with other RNA chains is not always required. Even if the ground state is perfectly ordered, with no discernible disordered regions, phase separation does take place, as was shown in experiments (8, 10) and confirmed using simulations (13) for repeat RNA sequences. 3) Because RNA is a polyanion, it stands to reason that cations could be efficacious in promoting phase separation by forming bridging interactions across RNA chains, although the mechanism of how this could be achieved is unknown (14).

Regardless of the scenarios, the driving force for aggregate formation must involve favorable free energy of association in order to compensate for the entropy loss due to confinement of the chains in droplets that, besides restricting exploration of all allowed conformations, also inhibits both translational and rotational motions. The free energy of association leads to an increase in the number of intermolecular WC base pair formation per chain relative to the monomer in isolation as well as enhanced entropy (13, 15) in the increased ways in which the base pairs can be paired in a condensate containing many chains. These arguments raise the following question: Could one anticipate the propensity of RNA chains to self-associate using the properties of RNA in the monomeric form?

Author contributions: H.M. and D.T. designed research; H.M., H.T.N., N.H., and D.T. performed research; D.T. contributed new reagents/analytic tools; H.M., H.T.N., N.H., and D.T. analyzed data; and H.M., H.T.N., N.H., and D.T. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. G.H. is a guest editor invited by the Editorial Board.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1] To whom correspondence may be addressed. Email: dave.thirumalai@gmail.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2301409120/-/DCSupplemental.

Published June 5, 2023.

Downloaded from https://www.pnas.org by UNIVERSITY OF TEXAS AT AUSTIN SER & ELEC RES DEPT PCL 2 302 on June 16, 2023 from IP address 128.62.216.51.

Downloaded from https://www.pnas.org by UNIVERSITY OF TEXAS AT AUSTIN SER & ELEC RES DEPT PCL 2 302 on June 16, 2023 from IP address 128.62.216.51.

We answer this question in the affirmative for low-complexity repeat RNA sequences, although the findings may be general.

Trinucleotide repeats, ubiquitous in human transcriptomes, are found in both the untranslated and translated regions of the human genome (16). A number of neurological and neuromuscular diseases, such as Huntington disease (HD), muscular dystrophy, and amyotrophic lateral sclerosis (ALS), are related to the expansion of repeat sequences that exceed a critical length (17). For instance, in healthy humans, the repeat length of cytosine–adenine–guanosine (CAG) Huntington genes varies from 16 to 20. In contrast, the length of CAG, resulting in HD, exceeds a critical value (∼35) (18). In an insightful paper, Jain and Vale (JV) (8) showed that $(CAG)_n$ polymers form biomolecular condensates provided n exceeds a critical number, which is close to the value required for the onset of the diseases. Here, we provide a quantitative description of the folding landscapes of $(CAG)_n$ monomers, as a function of n in monovalent salt solutions, in order to determine the nature of states that drives self-association between RNA chains.

We developed a coarse-grained (CG) model that is sufficiently accurate to simulate $(CAG)_n$ for arbitrary n. This genre of models has been proven to be useful in simulating the folding thermodynamics and kinetics in a variety of RNA molecules (19–23), with virtually no restriction on n or sequence diversity. Following our recent study (13), we used the Self-Organized Polymer (SOP) RNA model in which each nucleotide is represented by a single interaction site (24). Using the SOP model, we first investigated the structure and thermodynamics of $(CAG)_n$ as a function of n and monovalent salt concentration. The model predicts the melting temperatures that are in good agreement with available experiments. The populations of hairpins with perfectly aligned strands (PH) and slipped hairpins (SH), containing one or more overhangs, depend on whether n is even or odd and the concentration of monovalent salt, C_s. The ground state (GS) for odd (even) n, is the SH (PH). The slipped end in the SH is the source of multivalent interactions between the RNA polymers, which promote association of $(CAG)_n$. The free-energy gap, ΔG_S, between the GS and the excited state with (at least) one slip is a predictor of the propensity to form dimers and possibly the high-density droplets. In accord with this proposal, we show that dimer formation rates in $(CAG)_n$ for n = 29 and 31 are comparable and are greater than for n = 30. Our work suggests that ΔG_S between the GS and an excited state, with a minimum of one CAG overhang, is a good indicator of the propensity for self-association.

Results

Calculated Melting Temperatures for Small *n* Agree with Experiments. We first validated the SOP model simulations by comparing the calculated melting temperatures (T_Ms) with experiments. In the ground states of $(CAG)_n$, nucleotides G and C form Watson–Crick base pair with three hydrogen bonds. The only unknown parameter in the SOP model is the hydrogen bond strength u^0_{hb} (*SI Appendix*, Eq. S3) whose value is determined, as described in *SI Appendix*. We first performed folding simulations for the sequences, $G(CAG)_nC$ (n = 5, 6, and 7) at the monovalent salt concentration, C_s = 1 M. For these sequences, the experimental melting temperatures at C_s = 1 M (25) have been measured, which we used to extract the value of u^0_{hb} for use in all the simulations (*SI Appendix*, Fig. S1).

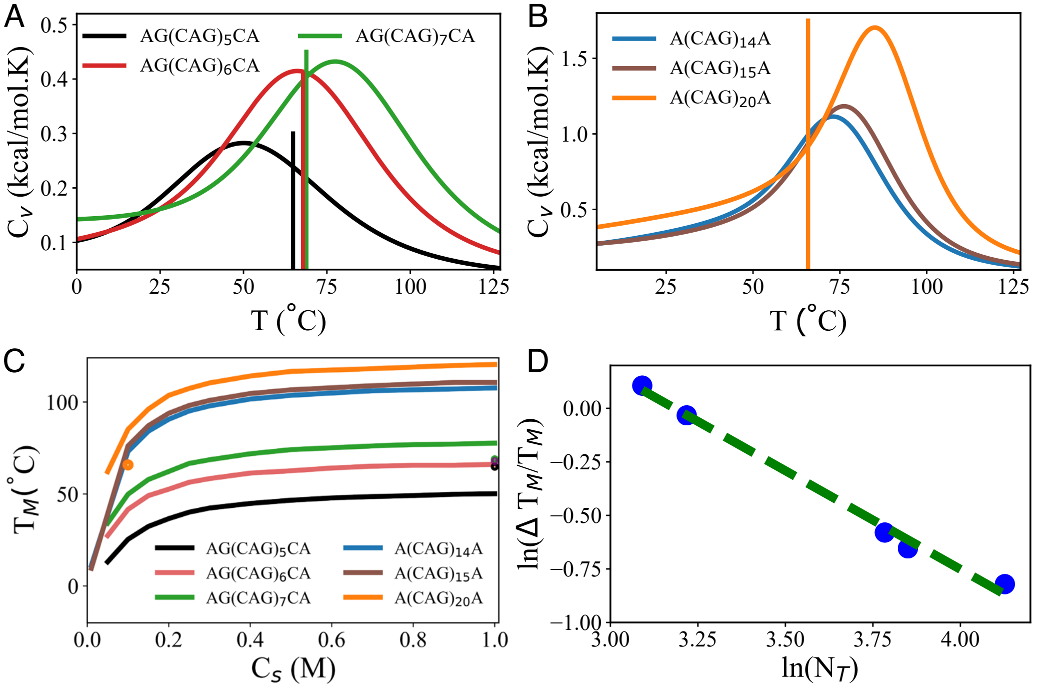

Fig. 1. Thermodynamics of $(CAG)_n$ sequences: (*A*) Heat capacity, C_V, as a function of temperature, T, for $G(CAG)_5C$ (black), $G(CAG)_6C$ (red), and $G(CAG)_7C$ (green) lines. The solid lines correspond to the experimental melting temperatures, T_Ms. (*B*) Same as (*A*) except that the plots are for $(CAG)_{14}$, $(CAG)_{15}$, and $(CAG)_{20}$ are in blue, cyan, and magenta lines, respectively. The experimental value for T_M for $(CAG)_{20}$ is shown as a solid line. (*C*) Predictions of T_Ms as a function of C_S for $G(CAG)_nC$ for various n that are indicated in the plot. The symbols are the measured values. (*D*) Log–Log plot of $\frac{\Delta T_M}{T_M}$ (ΔT_M is the full width at half maximum in $C_V(T)$) as a function of the number of nucleotides, N_T. The slope of the line is ≈ 0.92, which is close to the expected value of unity based on thermodynamic considerations.

Downloaded from https://www.pnas.org by UNIVERSITY OF TEXAS AT AUSTIN SER & ELEC RES DEPT PCL 2 302 on June 16, 2023 from IP address 128.62.216.51.

For each u^0_{hb}, we performed simulations in the temperature range $0° \leq T \leq 127\,°\mathrm{C}$ and calculated the heat capacity, $C_v(T) = \frac{\langle E^2\rangle - \langle E\rangle^2}{k_B T^2}$ (where $\langle E\rangle$ is the average of the potential energy, $\langle E^2\rangle$ is the associated mean square value), as a function of T. The location of the maximum in $C_v(T)$ is the melting temperature, T_M. The calculated T_M values, with $u^0_{hb} = -2.0$ kcal/mol, are 49 °C, 68 °C, and 77 °C for G(CAG)$_5$C, G(CAG)$_6$C, and G(CAG)$_7$C, respectively. The corresponding experimental values are ≈ 65 °C, 68 °C, and, 69 °C, respectively. The fair agreement for T_Ms between experiments and simulations (Fig. 1*A*) shows that the model, with a single parameter u^0_{hb}, is a reasonable starting point for predicting the outcomes for longer repeat lengths. In the simulations of all other sequences, we fixed $u^0_{hb} = -2.0$ kcal/mol.

Melting Temperatures of (CAG)$_n$ as a Function of Salt Concentration. We then simulated the folding of six (CAG)$_n$ (with $5 \leq n \leq 20$) as a function of C_s. In addition to n = 5, 6, and 7 (Fig. 1*A*), we also calculated C_v as a function of temperature at $C_s = 0.1$ M for (CAG)$_n$ with $n = 14$, 15, and 20 (Fig. 1*B*). The calculated T_M (≈ 81 °C) for (CAG)$_{20}$ at $C_s = 0.1$ M is in reasonable agreement with the experimental value (≈ 66 °C) (26). In principle, better agreement could be obtained by adjusting u^0_{hb}. Because our goal is to find general principles that govern the propensity of RNA sequence to phase-separate, we refrained from carrying out this exercise.

The C_s-dependent melting temperatures change substantially up to $C_s = 0.2$ M. As C_s increases further, the T_M increases almost linearly up to $C_s = 1$ M (Fig. 1*C* and *SI Appendix*, Fig. S2). The calculated melting temperatures for G(CAG)$_5$C, G(CAG)$_6$C, and G(CAG)$_7$C agree well with the experimental values determined by UV-absorbance measurements done at 1M monovalent (Na$^+$) salt concentration. The value of T_M for (CAG)$_{20}$ at C_s = 0.1M obtained using simulations ($T_M \approx 86\,°\mathrm{C}$) differs by about 20 °C (Fig. 1B) from experimental ($T_M \approx 66\,°\mathrm{C}$), as reported in Table 3 in ref. 26).

There could be two reasons for the discrepancy: 1) Experimental melting curves were measured using the standard UV absorbance (26) whose relation to the theoretical calculations based on energy fluctuations is unclear. 2) It is possible that the Debye–Huckel approximation used in the simulations is not accurate at low Na$^+$ or K$^+$ concentrations.

Finite Size Effects on T_M. As n increases, the width of the heat capacity curves decreases, which indicates that the first-order transitions are sharper (Fig. 1 *A* and *B*). From general thermodynamic considerations, it can be shown that the ratio $\frac{\Delta T_M}{T_M} \sim \frac{1}{N_T}$ where N_T is the total number of nucleotides. Previously, we have shown that this relation holds for proteins (27). Fig. 1*D* shows that for the (CAG)$_n$, the scaling holds perfectly.

Populations of Perfect Hairpin (PH) and Slipped Hairpin (SH) Vary for Even and Odd n. To understand the propensities to form condensates, we characterized the ensemble of structures explored by (CAG)$_n$ as a function of n. The (CAG)$_n$ sequences adopt multiple stem-loop or hairpin structures depending on the arrangement of the base pairs (bp) in the stem region. We characterized the hairpin structures using the order parameter, Q_{HP} (Eq. **1**), which quantifies the deviations in the arrangement of the bps of a given structure from a perfectly aligned hairpin (PH) (Fig. 2 *C* and *D*). An ensemble of conformations with $Q_{HP} = 0$ implies formation of bps that are identical to a PH with no mismatches, except the unavoidable A–A mismatch. The set of bps, S_{bp}, in a given conformation can be expressed using, $S_{bp} = \ (i, j) : i + j = N_T + 1$, where i and j are the indices of the nucleotides that form a base pair, and N_T is the total number of nucleotides in the sequence. In contrast to perfect hairpins, $Q_{HP} = m$ ($m > 0$), where the positive integer m is a measure of strand slippage, which can occur either at the 5′ or the 3′ ends. Such structures represent deviations from PH.

Conformations, with fractional Q_{HP}, typically have one or more bulges in the stem (*SI Appendix*, Fig. S8 *C*) region. The classification using Q_{HP} allows us to quantitatively represent the folding landscape of repeat RNAs in terms of the ground and excited states of the chain. We show below that the free-energy spectrum of states gives quantitative insights into the propensity for the RNA chains to self-associate.

We also calculated the probability distributions, $P(Q_{HP})$s, for the six sequences at C_s = 100 mM and temperature = 37 °C (Fig. 2). The most populated structure (GS) for (CAG)$_n$ depends on whether n is odd ($n = 2k + 1$) or even ($n = 2k$). For $n = 2k$, the state with the highest population is a PH with $Q_{HP} = 0$ (Fig. 2*A*), whereas for $n = 2k + 1$, the value of $Q_{HP} = 1$ in the ground state (Fig. 2*B*). It is worth noting that the GS population decreases as n increases regardless of whether n is odd or even. Interestingly, for even n, the population of the excited state with $m = 2$ (two slippage units) is modestly higher than with $m = 1$. For both $m = 1$ and $m = 2$, the number of base pairs is the same in the stem. However, formation of the hairpin-like structure with $m = 1$ requires a loop with seven nucleotides, whereas the excited state with $m = 2$ can be accommodated by the more stable tetraloop.

In addition to the GS, there are states with noninteger values of Q_{HP} that are also populated, albeit with smaller probabilities. The populations of such structures are greater if n is odd compared to even n (Fig. 2 *A* and *B*). The accessibility of such low free-energy structures for (CAG)$_{31}$ makes aggregation propensity greater than in (CAG)$_{30}$.

Our findings that (CAG)$_n$ with an odd (even) number of repeats form hairpin with slippage (no slippage) in strands are in accord with experiments (28) on single-stranded (CAG)$_n$ DNA and (CAG)$_{17}$ RNA (29). The GS configuration obtained for odd- or even-numbered repeats is also consistent with the prediction from the RNAfold web server generated at T = 37 °C and C_s = 1 M. Experiments (30) probing the mechanism for the conversion of (CTG)$_n$ DNA hairpins between blunt end ($Q_{HP} = 0$) and overhang (similar to $Q_{HP} = 1$) configurations suggest that the conversion occurs through the propagation of a bulge ($0 < Q_{HP} < 1$) in the stem.

Factors Contributing to the Stability of the Ground States. Alternation in the ground state structure in going from $n = 2k$ to $2k + 1$ may be explained using the stability of both the loop and stem regions. Representative hairpin structures for even- and odd-numbered repeats are given in Fig. 2 *C* and *D*, respectively (*SI Appendix*, Fig. S8). For $n = 2k$, the GS has $Q_{HP} = 0$ because such conformations accommodate the maximum number of bps, ($N^{max}_{bp} = k - 1$) in the stem region. In addition to $Q_{HP} = 0$, the hairpins with $Q_{HP} \leq 1$ also have the $N^{max}_{bp} = k - 1$ provided $n = 2k + 1$. As argued above, the preference for $Q_{HP} = 1$ over other structures is attributed to the stability of the loop region. All other conformations with $0 < Q_{HP} < 1$ contain at least one bulge in the stem (*SI Appendix*, Fig. S8 for examples). Electrostatic repulsion generated between the negatively charged phosphates of the unpaired nucleotides is a

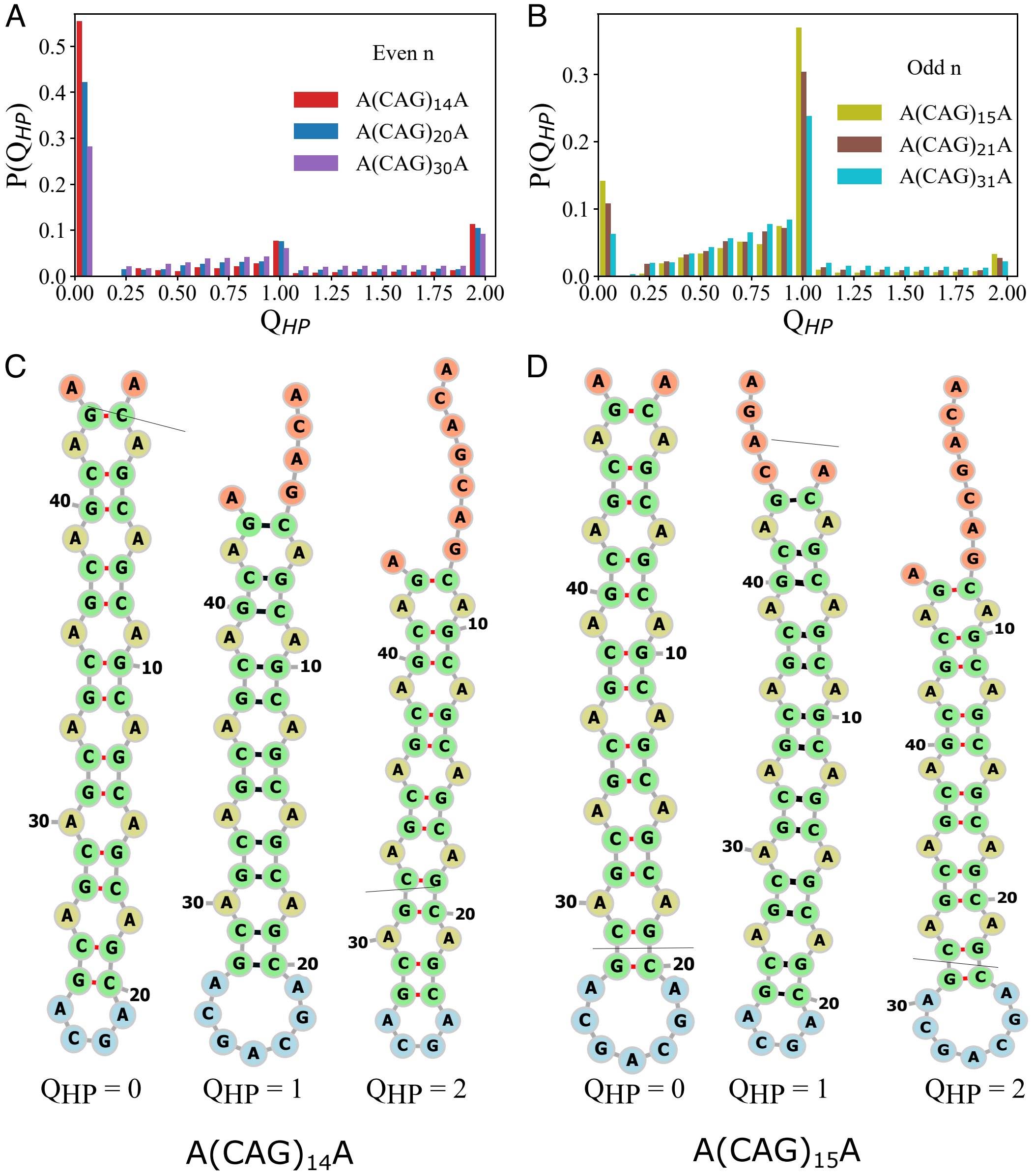

Fig. 2. Characterizing the hairpin-like structures: (*A*) Distribution, $P(Q_{HP})$, of the order parameter Q_{HP}, for (CAG)$_{14}$, (CAG)$_{20}$, and (CAG)$_{30}$. The maximum in $P(Q_{HP})$ is at $Q_{HP} = 0$, implying that the ground state is a PH. (*B*) Same as (*A*), except that $P(Q_{HP})$s are shown for odd values of *n*. The ground states have one unit of slippage, resulting in $Q_{HP} = 1$. (*C*) Representative hairpin structures with different Q_{HP} for (CAG)$_{14}$ as an example. (*D*) The most populated states along with the Q_{HP} values for (CAG)$_{15}$. Structures with fractional Q_{HP} are displayed in *SI Appendix*.

minimum for $n = 2k + 1$ if the sequence is in the Q_{HP} = 1 state. Thus, loop entropy and electrostatic interactions determine the ground states of the low-complexity sequences.

(CAG)$_{29}$ and (CAG)$_{31}$ Dimerize Faster than (CAG)$_{30}$. Our previous work (13) showed that homotypic interactions between (CAG)$_n$ chains results in the formation of droplets that coexist with monomers (or small oligomers) if *n* is large enough (typically around 30). Examination of the structural changes that occur when a single polymer chain adds on to a droplet revealed that a slipped state (initially with $m = 1$) forms first either from the $5'$ or the $3'$ end, exposing unsatisfied CG bases. Subsequently, the CG bases engage in complementary interactions with other chains in the droplet (13). This picture suggests that the ease of formation of the SH states could reveal the propensity to form droplets or duplexes or low-order oligomers. Therefore, we expect that the $\Delta G_S = G_m - G_{GS}$, the difference in free energy between the ground and the SH state, typically with one slippage ($m = 1$) at either end of the chain, should correlate with duplex formation. If the GS is SH, as is the case for (CAG)$_{31}$, then $\Delta G_S = 0$, which means that the GS is poised to interact with another RNA chain in order to initiate dimer formation. If this physical picture is correct, we expect that the time for forming a dimer should be smaller for (CAG)$_{31}$ compared to (CAG)$_{30}$ because ΔG_S for (CAG)$_{30}$ is larger than for (CAG)$_{31}$ (compare the spectra in Fig. 3 *A* and *B*).

From the calculated spectra for these two constructs (Fig.3 *A* and *B*), we expect that (CAG)$_{31}$ should dimerize faster than (CAG)$_{30}$. This expectation is based on the theory that the association rate depends on the population of states with exposed, at least partially, single-stranded regions. Because the ground state of (CAG)$_{31}$ is a slipped state, it should readily form inter-RNA interactions. In contrast, the slipped state in (CAG)$_{30}$ is a high free-energy state, with diminished population, relative to

Downloaded from https://www.pnas.org by UNIVERSITY OF TEXAS AT AUSTIN SER & ELEC RES DEPT PCL 2 302 on June 16, 2023 from IP address 128.62.216.51.

Downloaded from https://www.pnas.org by UNIVERSITY OF TEXAS AT AUSTIN SER & ELEC RES DEPT PCL 2 302 on June 16, 2023 from IP address 128.62.216.51.

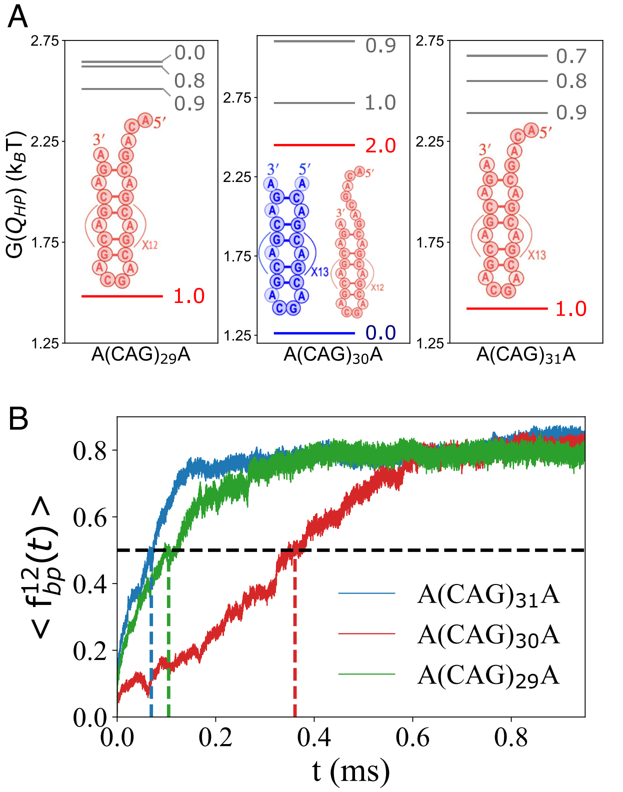

Fig. 3. Link between free-energy spectra and the kinetics of dimerization: (*A*) Free-energy spectra for (CAG)$_{29}$ (*Left*), (CAG)$_{30}$ (*Middle*), and (CAG)$_{31}$ (*Right*) computed at C_S = 0.1 M and T = 37 °C. The value of Q_{HP} in the ground state for the even sequence is zero, whereas it is unity for the odd sequences. (*B*) Time-dependent increase in the fraction of interchain base pairs, $f_{bp}^{12}(t)$ upon dimerization for (CAG)$_{29}$ (green), (CAG)$_{30}$ (red), and (CAG)$_{31}$ (blue). (CAG)$_{29}$ and (CAG)$_{31}$ dimers form nearly 5 times faster than the (CAG)$_{30}$ dimer, showing that the differences are not due to minor length difference, but it is an odd–even effect. Initially (at $t = 0$), the chains are in their ground state.

the ground state. Consequently, the dimerization rate should be less in (CAG)$_{30}$ than in (CAG)$_{31}$. To test this proposal, we simulated duplex formation (*SI Appendix* for details) for (CAG)$_{30}$ and (CAG)$_{31}$ by weakly constraining the polymers to remain within $R_0 = 40$ Å($\approx \langle R_g \rangle$, the monomer radius of gyration). The dimer simulations were initiated from the monomer ground states which is a PH (SH) for (CAG)$_{30}$ ((CAG)$_{31}$). To monitor dimer formation, we calculated the fraction of inter-molecular bp between the two chains (f_{bp}^{12}) as a function of time. We assumed that a dimer is formed if $f_{bp}^{12} > 0.5$. Out of the 100 trajectories, only in ≈ 40% of the trajectories the two chains adopted an antiparallel duplex (double-stranded RNA) structure. Fig. 3*B* shows the fraction, $f_{bp}^{12}(t)$, as a function of t. Based on the criterion that a dimer is formed if f_{bp}^{12} exceeds 0.5, we find that dimer formation is about 5 times slower in (CAG)$_{30}$ compared to (CAG)$_{31}$, thus qualitatively affirming the expectation based on the free-energy spectra in Fig. 3. The estimated time scale for duplex formation is a lower bound because the friction coefficient used in the simulations is much smaller than the value should be in water. Nevertheless, we believe that there is a link between the accessibility of slipped states to self-association kinetics should be fairly general.

Fig. 3*B* shows that the dimerization is faster in (CAG)$_{31}$ relative to (CAG)$_{30}$. Is the rate difference merely a consequence of length (there are three nucleotide difference between the two constructs) or is it an odd–even effect? In order to answer this question, we simulated dimer formation in (CAG)$_{29}$. Because the ground state for (CAG)$_{29}$ is a SH state (Left panel in Fig. 3*A*), we expect that the rate of dimerization should be comparable to (CAG)$_{31}$. If it is purely a length effect, we would predict that the time for dimer formation should be less in (CAG)$_{29}$ compared to (CAG)$_{30}$, which in turn should be less than (CAG)$_{31}$. In other words, the rates of dimerization should follow the trend, (CAG)$_{29}$ > (CAG)$_{30}$ > (CAG)$_{31}$. However, the simulations show (Fig. 3*B*) that the dimerization rates for (CAG)$_{29}$ and (CAG)$_{31}$ are similar, and both are faster compared to (CAG)$_{30}$. Thus, the difference between the dimerization rate in (CAG)$_{30}$ and (CAG)$_{31}$ is an odd–even effect that is reflected in the free-energy spectra (Fig. 3*A*), with length playing a subdominant role.

From the fraction of intermolecular base pair interaction between the two chains, f_{bp}^{12} (Fig. 3*B*), the absolute yield cannot be discerned. However, by counting the number of trajectories that form dimers in 1 ms, we find that the yields of the (CAG)$_{31}$ ∼ (CAG)$_{29}$ dimer is about 1.6 greater than (CAG)$_{30}$. Thus, the relationship between the free-energy gap and the dimerization rate is reflected in Fig. 3*B* and in the mutant simulations.

Dimerization Occurs Predominantly from the SH States. The conversion of hairpins to a dimer for (CAG)$_{31}$ occurs by two major pathways. Details of the structural transformations that occur along the predominant assembly pathway (Path I) are shown in Fig. 4 and *SI Appendix*, Fig. S11 for the minor pathway (Path II). In both the pathways, the hairpin transitions between different conformations before successful dimerization. The obligatory state prior to dimerization is the formation of the slipped state, either at the 5′ or 3′ terminus. Dimer formation occurs in 35 out of 40 trajectories (Path I). In all these trajectories, the slipped end of one of the hairpin starts interacting with the slipped end of the other chain. The structural transitions, measured using the fraction of intrachain and interchain base pairs as a function of time for one trajectory, show that the fraction of bp loss in each chain is compensated by gain in the interchain base pairs (Fig. 4*A*). The transition to the duplex state occurs over a short time window (given by the rectangle in black in Fig. 4*A*) during which bulk of the interchain base pairs (f_{bp}^{12}(t) in Fig. 4*A*) forms. Based on the mechanisms for the formation of dimers, we conclude that the propensity for the association of hairpins is higher if there is a slippage in the strand either at 5′ or 3′ termini. In other words, the free-energy gap ΔG_s, which is zero in (CAG)$_{31}$ is the determining factor in dimer formation. In the minority path II (5 out of 40 trajectories), the slipped end of one hairpin interacts with the loop region of the other hairpin and eventually forms the duplex structure (*SI Appendix* for details).

The structural transitions that occur in Path I are sketched in Fig. 4 *B*, *C*, *D*, and *E*. They show that increases in the slippage in one chain result in the loss of intrachain WC base pairs. But the loss is compensated by an increase in the interchain base pairs, which eventually leads to the perfect duplex formation (Fig. 4*E*). We showed in an earlier study (13) that a similar mechanism holds in condensate formation as well. The sequence of transitions that occurs in a single RNA chain (figure 6 in our previous study ref. 13) as it merges with the droplet is essentially the same as shown in Fig. 4. Thus, the driving force for condensate formation is the enhanced gain in forming not only many interchain CG base pairs but also the increase in

Downloaded from https://www.pnas.org by UNIVERSITY OF TEXAS AT AUSTIN SER & ELEC RES DEPT PCL 2 302 on June 16, 2023 from IP address 128.62.216.51.

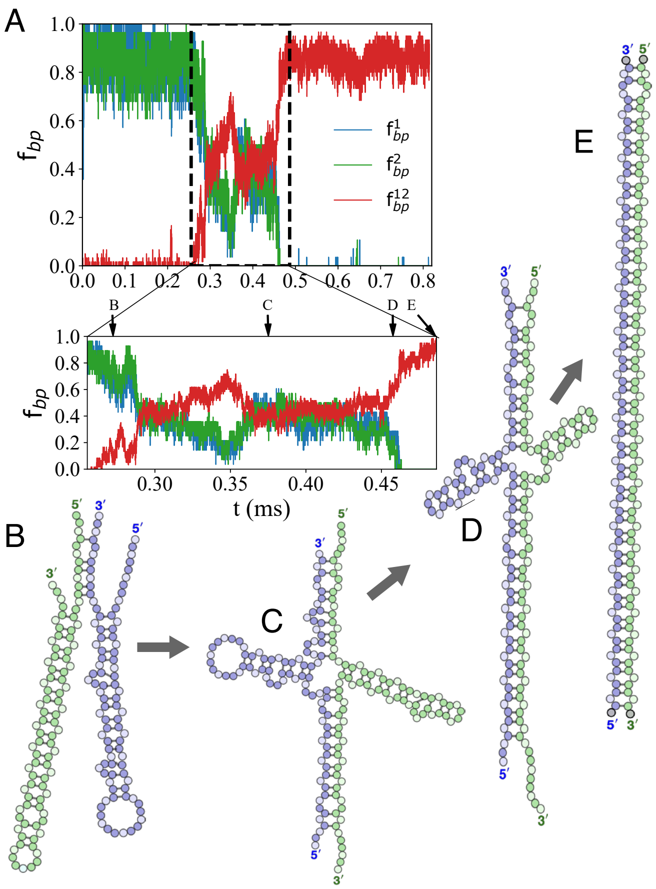

Fig. 4. Major dimerization pathway formation for $(CAG)_{31}$: (*A*) Time-dependent changes showing loss of fractions of intrachain base pairs ($f^1_{bp}(t)$ and $f^2_{bp}(t)$) and gain in interchain base pairs ($f^{12}_{bp})(t)$) in the dominant pathway. The initial structures for both the chains correspond to their ground states (Q_{HP} = 1). The value of C_S is 100 mM, and T = 37°. The panel below zooms in on the time range where the dimer forms. (*B–D*) Initial hairpin conformations first transition to the structures with a slipped ends at the termini. The hairpins with the slipped end start interacting with each other and eventually lead to the formation of a dimer through a sequence of transitions. Representative structures along the pathways are shown. Structural transitions in the minor pathway (*SI Appendix*, Fig. S11).

the number of ways in which such base pairs form, which is an entropic factor.

Effects of Mutations on Hairpin Structures and Dimer Formation. A corollary of the predicted inverse correlation ΔG_S and the propensity to dimerize is that suppression in the population of SH should reduce the yield of the dimer. To test this prediction, we probed the effects of mutations on the population of hairpin structures using variants of $(CAG)_{30}$ and $(CAG)_{31}$. The M1 mutant sequence is A (CCG)$(CAG)_{28}$(CGG)A, and the M2 sequence is A (CCG)$(CAG)_{29}$(CGG)A. We also considered other sequences (*SI Appendix*), but here, we focus on M1 and M2.

We calculated the populations ($P(Q_{HP})$ - shown in *SI Appendix*, Fig. S12) of various hairpin-like structures at C_s = 100 mM and T = 37 °C for M1 and M2 from which the free-energy spectra are readily calculated (Fig. 5 *A* and *B*). The ground state of M1 is a PH with $Q_{HP} = 0$, whereas the GS of M2 is a structure with a bulge but has no slipped CAG even though $Q_{HP} = 0.9$ is close to unity (Fig. 5). There are significant differences in $P(Q_{HP})$s between the WT and the mutants (Fig. 2 and *SI Appendix*, Fig. S12). For example, $P(Q_{HP} = 0)$ state in M1 increases significantly compared to the WT. In addition, ΔG_S for M1 is greater than in the WT. Comparison between the free-energy spectra for the WT and M2 also shows dramatic differences (compare Fig. 3*A* and Fig. 5*B*). The GS for the WT is a SH hairpin that is poised to form higher-order structures by interacting with other RNA chains. The values of ΔG_S in the WT and M2 are smaller compared with ΔG_S between the WT and M1. From the results in Figs. 4 and 5, we expect that the yield of the dimer in mutant sequences should be considerably less than in the WT sequences.

To illustrate the prediction that ΔG_S is the primary factor in determining the kinetics and yield of the duplex structures, we performed dimer simulations starting from the GS conformations for M1 and M2. Because the simulations are time consuming, we calculated the yield, instead of the rate of dimer formation. The percentage yield is calculated using, $P_D = \frac{N_D}{N_t} \times 100$ where N_D is the number of trajectories that reached the dimer state, and N_t is the total number of trajectories. We expect P_D to decrease as ΔG_S increases, which implies that P_D for the mutants must be greatly reduced relative to the corresponding WT sequences. There is a decrease in P_D, by at least one order of magnitude, in the mutant sequence compared to the WT sequences (Fig. 5 *C* and *D*).

Discussion

We created a coarse-grained model that allows for reasonably accurate calculations of the thermodynamic properties for monomer repeat RNA sequences as a function of monovalent salt concentration and sequence length. Because it is known (8) that a minimum length of repeat sequences is required for condensate formation, we also simulated $(CAG)_n$ for large n for both the wild type and several mutants, with the goal of linking their folding landscapes to the propensity of RNA polymers to self-associate. Two major findings that are likely to be valid for arbitrary RNA sequences emerge from our study. First, self-association is preceded by the formation of slipped hairpins even when most of the nucleotides are engaged in base pair formation. This is certainly the case when n is even. Second, there is an inverse correlation between the free-energy gap separating the ground state and the aggregation-prone slipped hairpin state and the rates and yields of dimer formation. It follows that structured RNA, with no exposed single strand, is unlikely to phase separate if the excited state is inaccessible.

Salt-Dependence of Folding Thermodynamics. The finding that the melting temperature, T_M, increases nearly linearly at small monovalent salt concentration, C_s, and more slowly beyond C_s = 0.2M (Fig. 1*C* and *SI Appendix*, Fig. S2) is amenable to experimental test. The free energy, $\Delta\Delta G(C_s) = \Delta G(C_s) - \Delta G(0.15M)$, decreases as $\Delta\Delta G(C_s) \sim -k\ln(C_s)$ for all $(CAG)_n$ (*SI Appendix*, Fig. S7).

Free-Energy Spectra as a Function of *n*. The differences between the ground state and the slipped hairpin free energies qualitatively explain the speedup in the dimer formation in $(CAG)_{31}$ compared to $(CAG)_{30}$. More precisely, it is the population of the SH that determines the aggregation rate. As n increases, we expect that ΔG_S should decrease, which means that for large n, the differences in the rates of self-association between RNA chains with odd and even n should decrease. A corollary of our finding is that mutations which increase ΔG_S should decrease dimerization rates, as illustrated in *SI Appendix*, Fig. S14.

Downloaded from https://www.pnas.org by UNIVERSITY OF TEXAS AT AUSTIN SER & ELEC RES DEPT PCL 2 302 on June 16, 2023 from IP address 128.62.216.51.

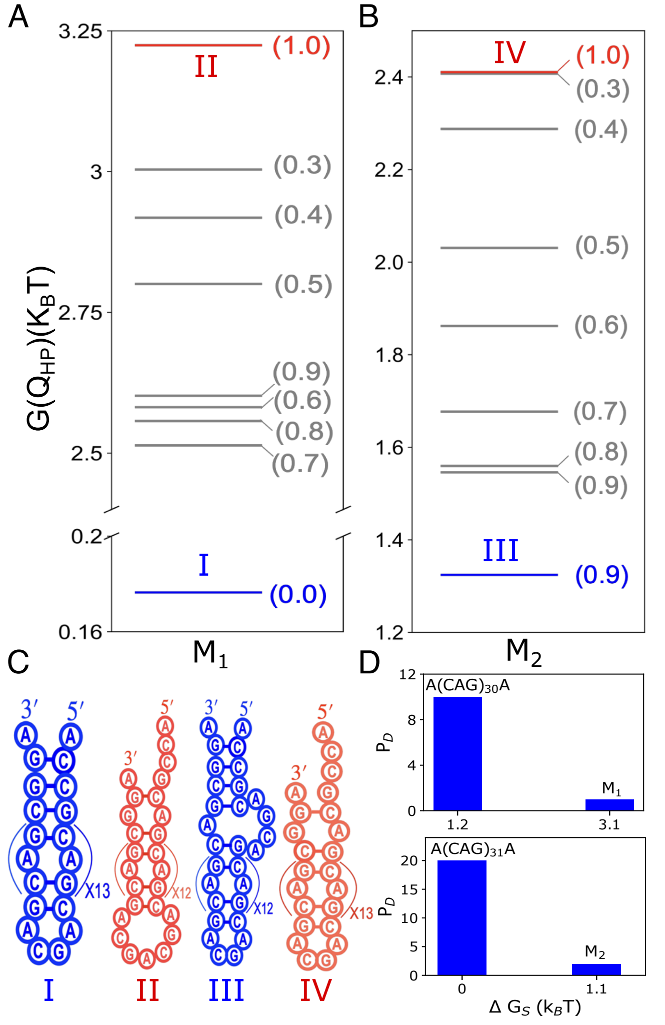

Fig. 5. Effect of mutations on dimer yields: (*A*) Free-energy spectrum of M1 mutant (A(CCG)(CAG)$_{28}$(CGG)A). The numbers represent the values of Q_{HP}. (*B*) Same as (*A*) except that the folding landscape is for M2 (A(CCG)(CAG)$_{29}$(CGG)A). The free energy between the ground state (PH) and the SH in M1 is higher than in M2. (*C*) Schematic structures of the relevant states (I, II, III, and IV) involved in dimer formation. (*D*) Dimer yields, expressed as percentage, for A(CAG)$_{30}$A and M1 mutant are in the *Top* panel, and the *Bottom* panels show A(CAG)$_{31}$A and M2. Note that the WT yield is higher in A(CAG)$_{31}$A than in A(CAG)$_{30}$A.

Finally, for a fixed *n* and identical sequence composition, the approximate dimer formation rate may be altered by changing the sequence. Thus, the exact sequence (10), in addition to *n*, salt concentration, and other environmental factors determine the tendency to self-associate.

Formation of a Slipped Hairpin is Obligatory for Self-association. The ground state of (CAG)$_{30}$ is a perfect hairpin, and yet it self-associates with another RNA chain to form a dimer. In order for this process to occur, both the RNA chains must transition to an excited state, resulting in the formation of SH, thus exposing one or more overhangs. The complementary sequences then could hybridize, thus nucleating the formation of the dimer. Our previous study showed that monomer addition to a droplet involves initial transition to a slipped state (figure 6 in ref. 13) that is followed by further unfolding that creates a large single-stranded conformation. The resulting excited states can form complementary base pairs with other chains in the droplet. Taken together, we ascertain that the formation of the SH state, regardless of the ground state, is a necessary condition for RNA chains to self-associate. A corollary is that if the formation of the SH is prevented, then droplets may not form or the time for self-association would greatly increase. This is the case in the mutant sequences in which either the dimerization time increases or the yield of the dimer is diminished substantially.

Sequence Diversity Should Inhibit Extensive RNA–RNA Interactions. Our findings raise a broader question: How are unwarranted interactions between various RNA chains, especially those that have to fold for functional reasons, in cells prevented? There are two possible reasons.

(1) Self-association between RNA chains requires exposure of single-stranded regions. Our findings suggest that at least one of the two RNA chains should be in the SH. Because $\Delta G_S/k_BT$ is not that large, the SH states are readily populated in the low-complexity sequences explored here. One measure of complexity is the sequence entropy given by $S(i) = -\sum_{k=1}^{4} p_k(i)\ln p_k(i)$, where $p_k(i)$ is the probability of finding a nucleotide of type k (A,T,G,C) at the i^{th} position in a multiple sequence alignment sense. For (CAG)$_n$, $S(i) \equiv 0$ for all i. In more complex RNA sequences ($S(i) > 0$—the maximum value is $ln4$), it is likely that ΔG_S is sufficiently large (exceeds $\approx 2k_BT$) that the population of states with exposed single strands is small. In order to demonstrate that this is indeed the case, we calculated the free-energy spectra for two hairpins, L8 and L10 (Fig. 6). The structure (labeled II in Fig.6*B*) with modest slippage (Q_{HP}=1.7) is separated from the ground state by $\approx 8k_BT$ in L8 (*Left* panel in Fig. 6*A*). In L10, the first appearance of a slipped state (labeled IV in Fig. 6*B*) in the free-energy spectrum (Fig. 6*A*) has Q_{HP} = 1.7 with a similar excitation free energy as in L8. *SI Appendix*, Fig. S15 shows that in conformations with Q_{HP} < 1.7, the 5′ and 3′ are close to each other without any overhangs. Therefore, the propensities of such structures to aggregate are diminished.

(2) RNA molecules, whose folding is needed for splicing (*Tetrahymena* ribozyme for example), misfold readily (31–34). However, the bases in the exposed regions in such misfolded states are often paired, although the structures are nonnative containing bulges, loops, and multiloops. Although it is known that in the ground states of large single-stranded RNA molecules the 5′ and 3′ are in proximity (35), much less is known about the structures of the excited states. It is more than likely (L8 and L10 being very simple examples) that the exposed single-stranded regions are not readily accessible in heterogeneous RNA sequences, which greatly suppresses the probability of association with other RNA molecules. Indeed, the excited state of P5abc sequence not only has low population (3%, which translates to $\Delta G_S \approx 3.5k_BT$) but the NMR structure (figure 3 in ref. 36) shows the absence of exposed single-stranded regions. Therefore, we surmise that in P5abc, conformations with any slippage, that could form complementary inter-RNA interactions, must far exceed $\approx 3.5k_BT$, which implies that self-association of P5abc chains is unlikely.

Of particular interest is the propensity to self-associate between mRNA molecules with identical sequences. Typically, such RNA molecules have 1,000 or more nucleotides. Although general theoretical arguments cannot be easily provided, the present work offers some important insights. If the 5′ and 3′ are spatially close,

Downloaded from https://www.pnas.org by UNIVERSITY OF TEXAS AT AUSTIN SER & ELEC RES DEPT PCL 2 302 on June 16, 2023 from IP address 128.62.216.51.

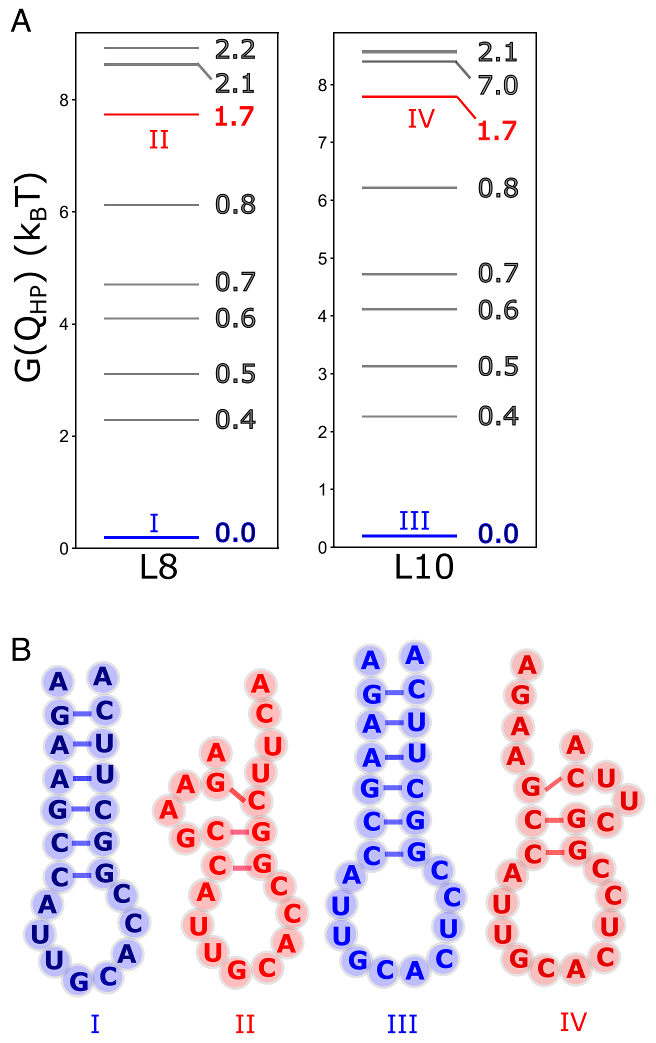

Fig. 6. Free-energy spectra and relevant structures for L8 and L10 hairpins: (*A*) Calculated spectra for hairpins with heterogeneity in the sequences. The free energies of the aggregation-prone states (in red) are separated from the ground states by about $8k_BT$, making the populations of such states vanishingly small. States with $Q_{HP} < 1.7$ do not have any slippage. (*B*) The ground state structures with 5′ and 3′ ends in proximity are in blue. The high free-energy states, with slipped nucleotides, are in red.

then the assembly-prone excited states would have extremely high free energies, which would suppress self-association. However, if there are fluctuations in one of the ends, thus exposing single-stranded strands, which by definition would be an excited state, then the assembly is likely. We believe that for generic diverse mRNA sequences, this scenario is unlikely. If there is a substantial probability of large fluctuations in either the 5′ or the 3′ end, then the local unfolding free energy would dictate the propensity to self-associate. Unlike for the low-complexity sequence, our work suggests that for biologically important mRNA sequences, extensive intermolecular association is unlikely, which supports previous experimental findings. (4).

Perhaps, the most important lesson from this work is that the determination of the propensity of RNA, with arbitrary sequence, requires elucidation of the entire ensemble of structures for a given specific environmental condition. This will require knowledge of excited states. Merely gazing at sequences, although useful, cannot yield much information about the propensity to self-associate.

Predicting Aggregation from Monomer Properties. Both in the context of protein folding and more recently in protein aggregation, there have been attempts to link the characteristics biophysical properties of monomers in determining the ease of self-association (37, 38). For instance, it was shown 30 y ago (39) that the efficiency of folding is determined by the proximity of the collapse (T_θ) and folding transition temperatures. More recently, it was argued (40) that T_θ and the temperature at which phase separation occurs are also correlated. These studies and the present work highlight the importance of investigating the phases of monomers in obtaining insights into the properties of multichain systems.

Conclusion

That the population of excited states determines protein aggregation propensity has been emphasized for a long time (41, 42). Conceptually, there are similarities between the monomer characteristics of protein monomers and their tendencies to aggregate. Indeed, this analogy has been used to explain RNA–RNA interactions, especially when the ribosome detaches from mRNA (43). The analogy to protein aggregation is most apt if the ground state of the protein is folded or native-like, which is the case in the amyloid formation in transthyretin and the associated mutants and aggregation of prion proteins (44, 45). In both these examples, aggregation must involve access to partially misfolded states (44, 46, 47), which would expose hydrophobic residues, and facilitate protein aggregation. In contrast, amyloid formation in peptides (Aβ_{42} or Fused in Sarcoma) aggregation-prone structures, with fibril-like character, are the excited states, while the ground states are disordered. Nevertheless, understanding self-association of RNA and proteins requires characterizing the entire free-energy landscape and not merely the ground states. Therefore, there is a great need for experimentally determining the free energies and structures of the excited states of RNA molecules, especially those that are involved in stress granule formation. This has been accomplished for proteins (48–50) and more recently for the P5abc domain (36).

Materials and Methods

We represent each nucleotide by a single site (13, 24). To account for ion condensation onto the highly charged RNA polyanion, we used a reduced value of $-Q$ on the phosphate groups with $0 < Q < 1$ (51, 52). Oosawa–Manning counterion condensation theory (53, 54) was used to calculate the value of Q. The total energy, E_{TOT}, of an RNA is the sum of bonded (E_B), hydrogen-bonded (E_{HB}), excluded volume (E_{EV}), and electrostatic energy (E_{el}). The details of the model and simulations are given in *SI Appendix*.

Structural Classification. In order to reveal the spectrum of free-energy states of the RNA sequences, we first calculated, $P(Q_{HP})$, the distribution of Q_{HP}, where Q_{HP} is an order parameter that measures the deviation in the arrangement of base pairs with respect to a perfectly aligned hairpin structure. We define Q_{HP} as,

$$Q_{HP} = \left(\frac{1}{N_{bp}}\sum_{i,j}^{N_{bp}}\left(\frac{(i+j-N_T-1)}{3}\right)^2\right)^{\frac{1}{2}}, \qquad [1]$$

where i and j label the nucleotides, N_T is the length of the sequence, and N_{bp} = number of base pairs in a given conformation. The nucleotide indices i and j, forming a base pair in a perfectly aligned hairpin, is related by $i+j = N_T+1$. A hairpin with $Q_{HP} = 0$ shows that the strands are perfectly aligned with respect to each other (the structures in Fig. 2*C*). Similarly, $Q_{HP} = 1$ corresponds to a slipped hairpin (SH) state (Fig. 2*D*). In the range $0 < Q_{HP} < 1$, the conformations contain bulges in the stem (*SI Appendix*, Fig. S8).

Downloaded from https://www.pnas.org by UNIVERSITY OF TEXAS AT AUSTIN SER & ELEC RES DEPT PCL 2 302 on June 16, 2023 from IP address 128.62.216.51.

Calculation of the Free-Energy Spectrum. In order to construct the free-energy spectrum, we first arranged the population of different hairpin structures, $P(Q_{HP})$, in descending order. The free energy of a given state, characterized by Q_{HP}, is calculated using,

$$G(Q_{HP}) = -k_BTln(P(Q_{HP})), \qquad [2]$$

where k_B is the Boltzmann constant and T is the absolute temperature. We constructed the free-energy spectrum by considering the structures for which $0 \lessapprox G(Q_{HP}) \lessapprox 4k_BT$. The structures with $G(Q_{HP}) > 4k_BT$ are not shown because their populations are negligible.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We are grateful to Tatjana Trcek for useful discussions. This work was supported by a grant from the NSF (CHE 19-00033) with additional support from the Welch Foundation (F-0019) through the Collie-Welch Chair.

Author affiliations: [a]Department of Chemistry, The University of Texas at Austin, Austin TX 78712; [b]School of Pharmacy, University of Nottingham, NG7 2rD, United Kingdom; and [c]Department of Physics, The University of Texas at Austin, Austin TX 78712

1. H. Eisenberg, G. Felsenfeld, Studies of temperature-dependent conformation and phase separation of polyriboadenylic acid solution at neutral ph. *J. Mol. Biol.* **30**, 17 (1967).
2. C. Roden, A. S. Gladfelter, RNA contributions to the form and function of biomolecular condensates. *Nat. Rev. Mol. Cell Biol.* **22**, 183–195 (2021).
3. B. Van Treeck *et al.*, RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2734–2739 (2018).
4. T. Trcek *et al.*, Sequence-independent self-assembly of germ granule mRNAs into homotypic clusters. *Mol. Cell* **78**, 941–950 (2020).
5. B. Van Treeck, R. Parker, Emerging roles for intermolecular RNA–RNA interactions in RNP assemblies. *Cell* **174**, 791–802 (2018).
6. W. M. Aumiller Jr, F. P. Cakmak, B. W. Davis, C. D. Keating, RNA-based coacervates as a model for membraneless organelles: Formation, properties, and interfacial liposome assembly. *Langmuir* **32**, 10042–10053 (2016).
7. T. Trcek *et al.*, Drosophila germ granules are structured and contain homotypic mRNA clusters. *Nat. Commun.* **6**, 7962 (2015).
8. A. Jain, R. D. Vale, RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
9. E. M. Langdon *et al.*, mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **360**, 922–927 (2018).
10. A. U. Isiktas, A. Eshov, S. Yang, J. U. Guo, Systematic generation and imaging of tandem repeats reveal multivalent base-pairing as a major determinant of RNA aggregation. *Cell Rep. Methods* **2**, 100334 (2022).
11. W. Ma, G. Zheng, W. Xie, C. Mayr, In vivo reconstitution finds multivalent RNA–RNA interactions as drivers of meshlike condensates. *eLife* **10**, e64252 (2021).
12. E. M. Langdon, A. S. Gladfelter, A new lens for RNA localization: Liquid–liquid phase separation. *Annu. Rev. Microbiol.* **72**, 255–271 (2018).
13. H. T. Nguyen, N. Hori, D. Thirumalai, Condensates in RNA repeat sequences are heterogeneously organized and exhibit reptation dynamics. *Nat. Chem.* **14**, 775–785 (2022).
14. P. L. Onuchic, A. N. Milin, I. Alshareedah, A. A. Deniz, P. R. Banerjee, Divalent cations can control a switch-like behavior in heterotypic and homotypic RNA coacervates. *Sci. Rep.* **9**, 1–10 (2019).
15. O. Kimchi, E. M. King, M. P. Brenner, Uncovering the mechanism for aggregation in repeat expanded RNA reveals a reentrant transition. *Nat. Commun.* **14**, 332 (2023).
16. W. J. Krzyzosiak *et al.*, Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.* **40**, 11–26 (2012).
17. C. Everett, N. Wood, Trinucleotide repeats and neurodegenerative disease. *Brain* **127**, 2385–2405 (2004).
18. P. McColgan, T. S.J., Huntington's disease: A clinical review. *Eur. J. Neurol* **25**, 24–34 (2018).
19. C. Hyeon, D. Thirumalai, Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6789–6794 (2005).
20. C. Hyeon, D. Thirumalai, Multiple probes are required to explore and control the rugged energy landscape of RNA hairpins. *J. Am. Chem. Soc.* **130**, 1538–1539 (2008).
21. N. A. Denesyuk, D. Thirumalai, Crowding promotes the switch from hairpin to pseudoknot conformation in human telomerase RNA. *J. Am. Chem. Soc.* **133**, 11858–11861 (2011).
22. N. A. Denesyuk, D. Thirumalai, How do metal ions direct ribozyme folding? *Nat. Chem.* **7**, 793–801 (2015).
23. N. Hori, N. A. Denesyuk, D. Thirumalai, Shape changes and cooperativity in the folding of the central domain of the 16S ribosomal RNA. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020837118 (2021).
24. C. Hyeon, R. I. Dima, D. Thirumalai, Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure* **14**, 1633–1645 (2006).
25. M. Broda, E. Kierzek, Z. Gdaniec, T. Kulinski, R. Kierzek, Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry* **44**, 10873–10882 (2005).
26. K. Sobczak *et al.*, Structural diversity of triplet repeat RNAs. *J. Biol. Chem.* **285**, 12755–12764 (2010).
27. Li. Mai Suan Klimov, K. Dmitri, D. Thirumalai, Finite size effects on thermal denaturation of globular proteins. *Phys. Rev. Lett.* **93**, 268107 (2004).
28. P. Xu, F. Pan, C. Roland, C. Sagui, K. Weninger, Dynamics of strand slippage in DNA hairpins formed by CAG repeats: Roles of sequence parity and trinucleotide interrupts. *Nucleic Acids Res.* **48**, 2232–2245 (2020).
29. K. Sobczak, M. de Mezer, G. Michlewski, J. Krol, W. J. Krzyzosiak, RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.* **31**, 5469–5482 (2003).
30. C. W. Ni, Y. J. Wei, Y. I. Shen, I. R. Lee, Long-range hairpin slippage reconfiguration dynamics in trinucleotide repeat sequences. *J. Phys. Chem. Lett.* **10**, 3985–3990 (2019).
31. D. Thirumalai, S. A. Woodson, Kinetics of folding of proteins and RNA. *Acc. Chem. Res.* **29**, 433–439 (1996).
32. D. K. Treiber, J. R. Williamson, Exposing the kinetic traps in RNA folding. *Curr. Opinion Struct. Biol.* **9**, 339–345 (1999).
33. S. Woodson, Recent insights on RNA folding mechanisms from catalytic RNA. *Cell. Mol. Life Sci.* **57**, 796–808 (2000).
34. D. K. Treiber, J. R. Williamson, Beyond kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* **11**, 309–314 (2001).
35. A. M. Yoffe, P. Prinsen, W. M. Gelbart, A. Ben-Shaul, The ends of a large RNA molecule are necessarily close. *Nucleic Acids Res.* **39**, 292–299 (2011).
36. Yi. Xue *et al.*, Visualizing the formation of an RNA folding intermediate through a fast highly modular secondary structure switch. *Nat. Commun.* **7**, ncomms11768 (2016).
37. B. Tarus, J. E. Straub, D. Thirumalai, Dynamics of Asp23-Lys28 salt-bridge formation in Aβ(10–35) monomers. *J. Am. Chem. Soc.* **128**, 16159–16168 (2006).
38. D. Chakraborty, J. Straub, D. Thirumalai, Energy landscapes of Aβ monomers are sculpted in accordance with Ostwald's rule of stages. *Sci. Adv.* **9**, eadd6921 (2023).
39. C. Camacho, D. Thirumalai, Kinetics and thermodynamics of folding in model proteins. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6369–6372 (1993).
40. G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, J. Mittal, Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9929–9934 (2018).
41. D. Thirumalai, D. Klimov, R. Dima, Emerging ideas on the molecular basis of protein and peptide aggregation. *Curr. Opin. Struct. Biol.* **13**, 146–159 (2003).
42. D. Chakraborty, J. E. Straub, D. Thirumalai, Differences in the free energies between the excited states of Aβ_{40} and Aβ_{42} monomers encode their aggregation propensities. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 19926–19937 (2020).
43. Nina Ripin, Roy Parker, Are stress granules the RNA analogs of misfolded protein aggregates? *RNA* **28**, 67–75 (2022).
44. R. I. Dima, D. Thirumalai, Probing the instabilities in the dynamics of helical fragments from mouse PrP^C. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15335–15340 (2004).
45. R. Moulick, R. Das, J. Udgaonkar, Partially unfolded forms of the prion protein populated under misfolding-promoting conditions. *J. Biol. Chem.* **290**, 25227–25240 (2015).
46. W. Colon, J. Kelly, Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry* **31**, 8654–8660 (1992).
47. P. Hammarstrom, R. L. Wiseman, E. Powers, J. Kelly, Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science* **299**, 713–716 (2003).
48. C. Tang, C. Schwieters, G. Clore, Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078–1082 (2007).
49. G. Bouvignies *et al.*, Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114 (2013).
50. P. Neudecker *et al.*, Structure of an intermediate state in protein folding and aggregation. *Science* **336**, 362–366 (2012).
51. N. A. Denesyuk, D. Thirumalai, Coarse-grained model for predicting RNA folding thermodynamics. *J. Phys. Chem. B* **117**, 4901–4911 (2013).
52. N. A. Denesyuk, N. Hori, D. Thirumalai, Molecular simulations of ion effects on the thermodynamics of RNA folding. *J. Phys. Chem. B.* **122**, 11860–11867 (2018).
53. F. Oosawa, *Polyelectrolytes* (New York, 1971).
54. G. S. Manning, Limiting laws and counterion condensation in polyelectrolyte solutions. I. Colligative properties. *J. Chem. Phys.* **51**, 924–933 (1969).